

After the acceptance of the manuscript modified by the authors according to the reviewers' comments the article has been published in Journal of Cereal Science:

Malegori, C., Grassi, S., Ohm, J. B., Anderson, J., & Marti, A. (2018). GlutoPeak profile analysis for wheat classification: skipping the refinement process. Journal of cereal science, 79, 73-79.

<https://doi.org/10.1016/j.jcs.2017.09.005>.

GlutoPeak profile analysis for wheat classification: skipping the refinement process

Cristina Malegori^{1,‡,†}, Silvia Grassi^{1,†}, Jae-Bom Ohm², James Anderson³, Alessandra Marti^{1,4,*}

¹ Department of Food, Environmental, and Nutritional Sciences (DeFENS), University of Milan, via G. Celoria 2, 20133 Milan, Italy

² USDA-ARS, Red River Valley Agricultural Research Center, Cereal Crops Research Unit, Hard Spring and Durum Wheat Quality Lab., 1250 Bolley Dr. Harris Hall 214, NDSU, Fargo, ND 58108-6050, USA.

Cereal Crops Research Unit, Red River Valley Research Center, Fargo, ND 58105, USA.

³ Department of Agronomy and Plant Genetics, University of Minnesota, 1991 Upper Buford Circle, St. Paul, MN 55108

⁴ Department of Food Science and Nutrition, University of Minnesota Twin Cities, 1334 Eckles Avenue, St. Paul, MN 55108

[†] These authors contributed equally to this work.

[#] Current address: Department of Pharmacy, University of Genova, viale Cembrano, 4, 16148 Genova, Italy

* Corresponding author: alessandra.marti@unimi.it (Alessandra Marti)

KEYWORDS

GlutoPeak test; farinographic stability; whole grain flour; Principal Component Analysis; k-Nearest Neighbours classification

ABBREVIATIONS

AgrEn, Aggregation Energy; EnMT, Energy to Maximum Torque; EnPM, Energy after Maximum Torque; FS, farinographic stability; GPU, GlutoPeak Unit; MT, Maximum Torque; PCA, Principal Component Analysis; PMT, Peak Maximum Time.

1 **ABSTRACT**

2 The GlutoPeak test can predict wheat flour quality by measuring gluten aggregation properties in a
3 short time and using a small amount of sample; thus has usefulness along the entire wheat delivery
4 chain. However, no information on the suitability of this new test on whole grain flours is available.
5 In this contest a multivariate approach was used to assess the GlutoPeak test ability to predict dough
6 quality directly from whole grain flour. GlutoPeak test was performed on both refined and whole
7 grain flours (22 samples), obtaining both profiles and calculated indices, which were subjected to
8 data exploration through Principal Component Analysis. Results suggested a trend according to
9 farinographic stability values. Furthermore, k-Nearest Neighbours classification models were
10 developed using the GlutoPeak profiles to predict farinographic stability, leading to average
11 prediction ability of 81.8% for both refined and whole grain flour data. The present outcome
12 suggests the possibility of predicting farinographic stability by GlutoPeak test directly on whole grain
13 flour, thus skipping the refinement process.

1. INTRODUCTION

Common wheat (*Triticum aestivum* L.) is the most suitable raw material for the production of a large number of products, including bread, biscuits, cakes, and noodles. The functionality and versatility of this cereal is associated with the capacity of its storage proteins - gliadins and glutenins - to form a gluten network that exhibits viscoelastic properties. However, although each wheat flour can organize its storage proteins into a viscoelastic network, different classes of wheat are suited for different types of products to deliver certain functional attributes. Indeed, hard wheat varieties are preferred for bread or others leavened products, where a strong gluten network is required; whereas soft wheat is preferred for cookies and cakes, where a weak gluten network is desirable. Even in the same class of wheat (e.g. hard wheat), not all the varieties perform the same during processing. In this regard, many rheological tests have been proposed - and successfully used - to predict wheat flour behaviour during dough mixing (e.g. Farinograph, Mixograph, Mixolab), extension (e.g. Extensograph, Alveograph), leavening (e.g. Rheofermentometer) and baking (e.g. Mixolab) (Pagani, Marti, & Bottega, 2014). Most of these rheological approaches have to adapt with the changes and, thus, the new needs of the value chain of wheat, where: (i) breeders look for reliable methods to test the functional quality of wheat lines at early stages, with just a limited amount of sample, (ii) millers need fast and reliable methods for checking wheat quality at the receiving station, and (iii) the baking industry looks for suitable methods that could predict end product quality for production and product development (Marti, Augst, Cox & Koehler, 2015). In other words, the value chain of wheat is looking for a reliable test for the prediction of gluten functionality in a short time and using a small amount of sample. In this frame, the GlutoPeak test has been recently proposed for the evaluation of wheat flour quality based on gluten aggregation kinetics (Kaur Chandi & Seetharaman, 2012). Recent work has shown GlutoPeak potential for predicting gluten quality for noodle (Lu & Seetharaman, 2014) and pasta (Marti, Cecchini, D'Egidio,

38 Dreisoerner, & Pagani, 2014) production. Moreover, the GlutoPeak indices have been used for
39 predicting the conventional parameters related to dough mixing stability, extensibility, and tenacity
40 (Marti, Ulrici, Foca, Quaglia, & Pagani, 2015).

41 Even if rheological analyses have a multivariate nature, few studies dealing with grain and
42 cereal products treat the data with a chemometric approach (i.e. Pastor, Ačanski, Vujić, Bekavac,
43 Milovac, & Kravić, 2016; Banu & Aprodu, 2015; Vidya, Ravi, & Bhattacharya, 2013). At present,
44 Chemometrics is being more and more applied all along the food chain for general exploratory data
45 analysis (through Principal Component Analysis - PCA) (Giovenzana, Beghi, Malegori, Civelli, &
46 Guidetti, 2013), multivariate regression (such as Partial Least Square Regression - PLSR) (Malegori,
47 Marques, de Freitas, Pimentel, Pasquini, & Casiraghi, 2017) and classification (with discriminant or
48 one-class methods) (Oliveri & Downey 2012). In a previous study, Marti, Ulrici, Foca, Quaglia, &
49 Pagani (2015) applied PCA and PLSR to investigate the correlations existing between the
50 conventional wheat rheological parameters and the GlutoPeak indices. However, the mentioned
51 approach relied on the GlutoPeak indices calculation from a dedicated software, whereas the direct
52 extrapolation of relevant information from the GlutoPeak profile would reduce the calculation time,
53 overcome the software dependence and capture its multivariate nature.

54 To the best of our knowledge, the GlutoPeak test has been applied only on refined flours,
55 with the exception of the work done by Kaur-Chandi & Seetharaman (2012) with the aim of
56 optimizing the method. Thus, the potential use of the GlutoPeak test on whole grain flours has not
57 been explored yet, in view of describing wheat quality. Defining gluten quality directly on the wheat
58 kernels would be of great interest for both breeders and millers, since it would predict wheat quality
59 while skipping the refinement process, and thus saving time.

60 Taking into consideration the multivariate nature of rheological data, the aim of this study was to
61 investigate GlutoPeak profiles by chemometric techniques for gluten aggregation properties

prediction of both whole grain and refined flours. To the aim, a GlutoPeak data exploration step was performed through PCA, followed by the development of classification models for farinographic stability prediction by k-Nearest Neighbours algorithm (Vandeginste, Massart, Buydens, De Jong, Lewi, & Smeyers-Verbeke, 1998). Moreover, the possibility of using the GlutoPeak curve of whole grain flours for the assessment of the refined wheat flour quality category was also investigated, following the same approach proposed for refined wheat flour. This study would provide information whether it would be possible to predict dough properties from whole grain samples without the need for the refining process, thus allowing the classification of a sample according to end-product functional attributes immediately at raw material acceptance.

2. MATERIALS AND METHODS

2.1 Materials

A set of 22 hard red spring wheat varieties were cultivated in 2011 in Crookston (MN, US). Whole grain flours were obtained by grinding the seeds with a Cyclone Sample Mill (UDY Corp., Fort Collins, CO) equipped with a 0.25 mm screen. From the same source samples, refined flours were obtained by conditioning (to 15.5 g/100g moisture overnight) and milling the seeds with a Quadrumat Junior (C.W. Brabender Inc., South Hackensack, NJ, USA).

2.2 Farinograph Test

Grain samples were analyzed for mixing properties using the Farinograph-E (CW Brabender Instruments, Inc., South Hackensack, NJ, US) equipped with a 10 g mixing bowl (AACCI 54–21, 2000).

2.3 GlutoPeak Test

Gluten aggregation properties were measured using the GlutoPeak device (C.W. Brabender Inc., South Hackensack, NJ, USA), as reported by Kaur-Chandi & Seetharaman (2012). An aliquot of 8.5 g of sample (refined and whole grain flour) was dispersed in 9.5 g of 0.5 mol L⁻¹ CaCl₂, scaling both solvent and flour weight on a 14% flour moisture basis in order to keep the liquid-to-solid ratio

86 constant. Sample temperature was maintained at 34 °C by circulating water through the jacketed
87 sample cup. The paddle was set to rotate at 1900 rpm and the test was carried out for 7 min.

88 **2.4 Data processing**

89 *2.4.1 Index calculation*

90 The main indices automatically evaluated by the software provided with the GlutoPeak device
91 (Brabender GlutoPeak v. 2.1.2) are: *i)* the Peak Maximum Time (PMT, expressed in s), corresponding
92 to the time before torque falling off when gluten breaks down; *ii)* the Maximum Torque (MT,
93 expressed in GlutoPeak Unit - GPU), corresponding to the peak occurring due to gluten aggregation;
94 *iii)* aggregation energy (AgrEn, expressed in GlutoPeak Unit - GPU), corresponding to the area under
95 the curve between 15 s before and 15 s after MT. In addition, the following indices were calculated
96 using Microsoft Excel 2010 (Microsoft, Redmond, VA): *iv)* Energy to Maximum Torque (EnMT;
97 expressed in arbitrary units - AU) corresponding to the area of the curve from the beginning of the
98 test and MT; *v)* Energy after Maximum Torque (EnPM; expressed in arbitrary units - AU)
99 corresponding to the area of the curve from the beginning of the test and 15 s after MT.

100 *2.4.2 Multivariate approach*

101 The GlutoPeak profiles collected from the 22 wheat flour samples, as well as a dataset composed
102 by the calculated indices, were submitted to two separated PCA after data mean centering. The
103 same approach was followed for the data collected by analysing the related whole grain flours.
104 Beyond being a powerful tool for data visualization (Jolliffe, 1986), PCA provided an understanding
105 of the relationships between all the variables (i.e. variables which contribute similar information to
106 the model) and among variables and samples. Thus, it was possible to compare samples patterns in
107 each PCA model obtained with both GlutoPeak profile and calculated indices; in addition, the
108 variables' weight in the new defined plane was assessed.

109 **As a first attempt, linear classification methods (namely, linear discriminant analysis – LDA – on the**

principal component scores, and partial least square discriminant analysis – PLS-DA) were applied, but these techniques did not provide satisfactory results in terms of classification rate (data not shown). This indicated a non-linear method to be required. The K-nearest neighbors (K-NN) technique was chosen for the classification of both whole grain and refined flour, being one of the less prone to overfitting thanks to the simplicity of its algorithm, which does not include optimization of numerous parameters. The definition of the two “a priori” classes was based on the FS value, allowing to obtain two balanced classes in term of number of constituents. K-NN, without making any assumptions on the underlying data distribution, keeps all the training data (the 22 GlutoPeak profiles) and calculates the Euclidean distance of the data points in their metric space. In our case, using the 1-nearest neighbour classifier, the simpler nearest neighbour type classifier, the algorithm assigns each sample to the class of its closest neighbour in the feature space (Dudoit, Fridly, & Speed, 2002). The reliability of the classifier was tested by leave-one-out cross-validation due to the reduced dimension of the dataset. The error rate of each k-Nearest Neighbour classifier was expressed as prediction rate. Data processing was carried out by a Matlab R2016a (The MathWorks. Inc, USA) routine.

3. RESULTS AND DISCUSSION

3.1 Mixing properties and gluten aggregation kinetics

Table 1 reports the farinographic stability (FS) values of refined flours from hard red spring wheats. All the analysed flour samples exhibited a mixing profile typical of strong flours characterized by high FS (from 9 to 42.9 min for RF21 and RF17, respectively). Beside FS, flours showed high water absorption (ranging from 61.4 to 71 g/100g for RF10 and RF16, respectively; data not shown) and long development time (from 6.1 to 33.7 min for RF6 and RF9, respectively; data not shown), suggesting high quality gluten and good bread-making performance.

133 Dough stability index - expressing the length of time from the arrival of the maximum torque line to
134 500 BU to the departure of the maximum torque line from 500 BU - provides information on dough
135 resistance to prolonged mixing: the higher the FS, the stronger the gluten network. Flours with FS
136 greater than 16 min are considered the most suitable raw material for manufacturing highly
137 leavened products, which requires long fermentation time (Foca et al., 2007).

138 In Table 1 are also summarised the calculated indices from the GlutoPeak profiles of refined flours;
139 whereas the latter are presented in Fig. 1 together with the related whole grain samples profiles.

140 During the test, the sample slurry is subjected to intense mechanical action promoted by the speed
141 of the rotating element. The increase in torque corresponds to the formation of the gluten network.
142 After reaching a maximum value, the torque curve declines, due to prolonged mixing at high speed
143 which causes the breakage of gluten network (Marti, Cecchini, D'Egidio, Dreisoerner, & Pagani,
144 2014). Usually, hard wheat flours exhibit longer aggregation time (i.e. peak maximum time) and
145 higher maximum torque (i.e. peak torque) than flours of soft wheat cultivars (Lu & Seetharaman,
146 2014), while flours for wafers or batters show very much delayed peak formation and much lower
147 torque (data not shown). More recently, the area under the peak (i.e. energy to peak), which takes
148 into consideration both the indices, has been found suitable for predicting conventional parameters
149 related to dough strength (Marti, Ulrici, Foca, Quaglia, & Pagani, 2015). To the best of our
150 knowledge, this is the first study exploring the gluten aggregation profile of hard spring varieties.
151 Spring wheat samples showed a faster aggregation (i.e. lower PMT), higher MT, and a broader peak
152 than winter wheats (Marti, Augst, Cox & Koehler, 2015). The use of different versions of the software
153 makes the energy data somewhat difficult to compare with winter wheat. For spring wheat varieties
154 here considered, energy to peak ranges from 1485 to 3843 AU for RF18 and RF9 varieties,
155 respectively. A moderately correlation ($r = 0.59$; $p = 0.005$) was found between this index and FS.
156 The software also provides the aggregation energy, expressed in GPU, which is the area under the

157 curve between 15 s before and 15 s after the peak. However, for this set of samples, this parameter
158 was not significantly correlated with FS. Indeed, looking at the GlutoPeak profile, most of the curves
159 exhibited a very broad peak, instead of a sharp one (Fig. 1). Thus, for this kind of shape, the area
160 under the curve till 15 s after peak (i.e. EnPM) should be considered, being moderately correlated
161 ($r = 0.62$; $p = 0.0019$) with FS.

162 Whole grain flour showed a rapid build up in consistency to a sharply defined peak followed by a
163 rapid break down, while the related refined flour showed a much slower build up in dough
164 consistency and a relatively more time to achieve peak consistency (Fig. 1).

165 Concerning whole grain profiles, a high MT and a low PMT are observed and resulted in an overall
166 low aggregation energy, suggesting a weak gluten network. A similar trend in gluten aggregation
167 profile has been found adding fiber-rich raw materials to refined wheat flour (Marti, Torri, Casiraghi,
168 Franzetti, Limbo, Morandin, Quaglia, & Pagani, 2014; Marti, Qiu, Schoenfuss, & Seetharaman, 2015).
169 This behaviour is likely due to gluten dilution, accounting for the decrease in energy and the fast
170 aggregation. On the other hand, the increase in MT could be related to either the ability of fiber to
171 absorb water or the higher kernel protein content compared to refined flours. Indeed, it has been
172 observed previously that increasing protein content of wheat flour led to an increase in torque
173 (Marti, Augst, Cox, & Koehler, 2015; Marti, Ulrici, Foca, Quaglia, & Pagani, 2015).

174 **3.2 Multivariate approach**

175 This work proposes an innovative approach for GlutoPeak data, which intends to highlight
176 interesting information without any index calculation. Such a multivariate data processing could be
177 of particular interest when considering whole grain analysis, for which a specific index value could
178 give a misleading information due to peculiar GlutoPeak profile generated and the lack of
179 background in the whole grain analysis by GlutoPeak test.

180 Until now, a chemometric approach was applied only to evaluate together all the discrete indices

181 traditionally calculated on the GlutoPeak curve (Marti, Ulrici, Foca, Quaglia, & Pagani, 2015); in the
182 present study, a further step was introduced, carrying out PCA directly on the profiles.

183 3.2.1 Refined flours: data exploration

184 Fig. 2 compares PCA results of the two approaches (discrete indices vs. whole curves): Fig. 2a and
185 Fig. 2b are, respectively, the score and the loading plot obtained considering the five most used
186 indices (PMT, MT, EnMT, EnPM and AgrEn); Fig. 2c and Fig. 2d are the score plot and the loading
187 plot of the profiles. The samples in both of the score plot are coloured according to a chromatic
188 scale that is proportional to the FS value, from blue – for low FS value – to red – for high
189 farinographic stability. The score plot calculated from the five indices matrix shows a trend
190 according to the FS values, indeed samples with low FS are arranged in the left-bottom quarter (dark
191 blue samples) whereas samples with FS values higher than 23.5 min are in the right-top quarter with
192 the exception of sample RF19 (Fig. 2a). Comparing the score plot, the evaluation of the whole profile
193 (Fig. 2c) allows better highlighting of a trend, from the left-bottom to the right-top corner, in the
194 orthogonal space; also in this case sample RF17 appears as an outlier. All the samples with FS higher
195 than 17.5 min have positive PC1 scores except sample RF17 (Figure 2c), suggesting a boundary
196 between two classes (FS>17.5 min and FS<17.5 min) to be used for further classification approaches.
197 A detailed analysis of loading plots, in comparison with the corresponding score plots, allows a
198 deeper understanding the relationships among variables and between variables and samples.
199 In particular, for the indices (Figure 2b), the loading plot shows associated discrete variables that
200 are positively correlated in terms of useful information (AgrEn and MT, EnMT and EnPM). It is
201 possible to notice that AgrEn and MT are positioned in the right-bottom quarter, thus they stretch
202 samples that have positive PC1 values and negative PC2 values in the score plot; whereas EnMT and
203 EnPM are in the right-top of the loading plot, contributing to split samples with high FS values (> 30
204 min) from the others. PMT shows a different behaviour as it is located in the left-top of the loading

205 plot, the same region where it is possible to find samples RF11, RF15 and RF18. Combining the
206 loading information and the fact that these samples have central FS values in respect to the analysed
207 set, it is plausible to believe that this index is not particularly correlated with sample grouping
208 according to FS characteristic.

209 The loading plot of the first two principal components along time (Fig. 2d) highlights the two most
210 important regions of the GlutoPeak profiles: around 75 s, loadings exhibit high positive values on
211 PC1 and negative values on PC2, whereas, around 150 s, an inverse trend is observed. By
212 comparison with the GlutoPeak profiles, it is possible to notice that the changes highlighted by PCA
213 loadings are linked to the average time of the beginning of gluten aggregation, i.e. the maximum
214 peak slope (75 s), and the average time before torque falling off when gluten breaks down (150 s).

215 Thanks to the joint interpretation of scores and loadings, it is possible to notice that the horizontal
216 shift along the time axis, characterizing the very broad peaks of flour profiles, is associated with
217 different FS values: peaks shifted towards longer stability time are associated with samples with FS
218 values higher than 17.5 min, while earlier peaks are associated to FS values lower than 17.5 min. In
219 other words, the joint analysis confirms what was already commented by the scores evaluation:
220 samples with FS values lower than 17.5 min are located in the right-bottom corner of the PC score
221 space, characterized by positive scores on PC1 and negative scores on PC2, which correspond to the
222 relative loading intensities of the peak at 75 s. On the contrary, samples located towards the left-
223 top corner are associated with a delayed peak (around 150 s), as evidenced by the loading trend.
224 Thus, it is possible to assess that PCA using the GlutoPeak profile gives thorough information about
225 samples considering their FS, bypassing the calculation of specific indices, which, in some case (i.e.
226 PMT) could be not that informative in respect to the studied property and doing the groundwork
227 for sample classification.

228 3.2.2 Whole grain: data exploration

229 The same approach tested for the GlutoPeak profiles of refined flour was performed to verify if it is
230 possible to extract relevant information about dough properties directly from whole grain samples
231 without a refining process. Therefore, PCA was carried out and the outcomes are presented in Fig.
232 3. In the score plot (Fig. 3a) a trend similar to that observed for refined flour samples can be
233 observed, although less evident and mainly ascribable to PC1. In particular, whole grain flour from
234 samples with FS values higher than 17.5 have negative PC1 scores, confirming the classes boundary
235 forecast by PCA on refined flour GlutoPeak profiles. Regarding the loading plot (Fig. 3b), only one
236 interesting peak could be highlighted, around 40 s. Indeed, the peak characterizing PC1 loadings
237 describes differences between samples linked to the high **torque** and the shift of the peak occurring
238 around 40 s in the GlutoPeak, also associated with torque intensity, suggesting that the higher the
239 peak for whole grain flours, the lower the FS value. Indeed, in whole grain samples, MT was
240 significantly correlated to bran content ($r = 0.62$; $p = 0.0019$), that negatively affect dough formation
241 and its strength during mixing, decreasing FS.

242 PCA confirmed to be a reliable method to explore the variance hidden in the data. In this case, the
243 main variables linked to samples differentiation were identified directly on the whole grain
244 GlutoPeak profile, allowing us to bypass calculation of indices that could lead to controversial
245 conclusion on the gluten aggregation properties of the final product.

246 **3.5 Refined flours and whole grain: classification**

247 On the basis of the outcomes obtained from the exploratory study, the final aim of this work was to
248 classify hard red spring wheat samples according to FS value using the whole GlutoPeak curves. To
249 perform a supervised classification, it is necessary to define a-priori classes using a FS cut-off; since
250 in the PCA score plots of both refined and whole grain flours it is identifiable a sample grouping
251 linked to PC1 values, two classes were defined setting as threshold 17.5 min of FS.

252 **In the choice of the best K-NN classification rule, several parameters were considered, including**

253 data pre-processing (raw data, standard normal variate transform, first- and second-order
254 derivatives, with the Savitzky-Golay method, third order polynomial, 11 datapoint window),
255 distance (Euclidean, Mahalanobis) and number of K neighbors (odd values from 1 to 7). All of the
256 combinations were evaluated, looking for the highest classification rate in fitting. The best
257 conditions, in term of classification rate in fitting, were found for raw data, Euclidean distance and
258 $K = 1$. The reliability of the method was tested by leave-one-out cross-validation expressing results
259 as misclassification matrix and related prediction rate (Table 2).

260 Although sample trends within the score plot were less evident for whole grain samples, comparable
261 results were obtained in term of average prediction ability (81.8%); membership attribution is more
262 balanced – in term of prediction rate of the two categories - using refined flour curves while, with
263 whole grain data, it is more effective at predicting samples with high FS values. In particular, the
264 classification model built for flour samples performed well in discriminating samples of both classes,
265 missing less than 21% of the samples no matter the considered class. The whole grain model
266 performed better in identifying samples with FS values higher than 17.5 min, indeed 13 samples out
267 of 14 were correctly recognized as member of class 1; whereas class 2 prediction ability was 62.5 %.

268 Sample 17, in both whole grain and refined flour models, resulted misclassified. Even in the PCA
269 results it behaved as an outlier, however it was not removed from the classification analysis as the
270 number of samples was already low.

271 **4. CONCLUSIONS**

272 The multivariate analysis of gluten aggregation kinetics of hard red spring wheat varieties
273 highlighted that the GlutoPeak test is able to classify wheat according to dough stability, i.e. the
274 farinographic index widely used for defining wheat end-uses according to gluten strength.

275 In addition, wheat classification can be obtained by analysing directly the GlutoPeak curves of the
276 whole grains flour, skipping the conventional milling process to obtain the refined flour.

277 The use of GlutoPeak test as a rapid small-scale test for wheat classification on whole grain flour
278 could have a tremendous impact on breeders and the milling industry because it requires only a
279 small sample and is characterized by rapidity, reliability, and little/absence of technical skill. At the
280 same time, the possibility of obtaining reliable results also on whole grain flours makes the
281 GlutoPeak the potential ideal test for evaluating large numbers of breeder's lines and wheat
282 varieties, thus facilitating quality evaluations at early stages of a breeding program and at the milling
283 receiving station, respectively.

284 Last but not least, instead of using the indices calculated by the software, the application of
285 multivariate analysis on the curve highlighted that the GlutoPeak profile might “hide” useful
286 information related to gluten strength.

287 Additional studies are required to confirm these findings on different type of wheat – including soft
288 wheat and hard winter wheat – and on a larger set of samples.

289 **References**

- 290 Banu, I., & Aprodu, I. (2015). Association of physicochemical with technological properties of
291 wheat. *International Journal of Food Science and Technology*, 50, 1644-1650.
- 292 Dudoit, S., Fridly, J., & Speed, P. (2002). Comparison of discrimination methods for the
293 classification of tumors using gene expression data. *Journal of the American Statistical Association*,
294 97, 77–87.
- 295 Foca, G., Ulrici, A., Corbellini, M., Pagani, M. A., Lucisano, M., Franchini, G. C., & Tassi, L.
296 (2007). Reproducibility of the Italian ISQ method for quality classification of bread wheats: An
297 evaluation by expert assessors. *Journal of the Science of Food and Agriculture*, 87, 839-846.
- 298 Giovenzana, V., Beghi, R., Malegori, C., Civelli, R., & Guidetti, R. (2013). Wavelength selection
299 with a view to a simplified handheld optical system to estimate grape ripeness. *American Journal of*
300 *Enology and Viticulture*, 65, 117-123.

301 Kaur-Chandi, G., & Seetharaman, K. (2012). Optimization of gluten peak tester: A statistical
 302 approach. *Journal of Food Quality*, 35, 69-75.

303 Lu, Z., & Seetharaman, K. (2014). Suitability of Ontario-grown hard and soft wheat flour
 304 blends for noodle making. *Cereal Chemistry*, 91, 482-488.

305 Malegori, C., Marques, E. J. N., de Freitas, S. T., Pimentel, M. F., Pasquini, C., & Casiraghi, E.
 306 (2017). Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the
 307 evaluation of acerola fruit quality, using PLS and SVM regression algorithms. *Talanta*, 165, 112-116.

308 Marti, A., Cecchini, C., D'Egidio, M. G., Dreisoerner, J., & Pagani, M. A. (2014).
 309 Characterization of durum wheat semolina by means of a rapid shear-based method. *Cereal*
 310 *Chemistry*, 91, 542-547.

311 Marti, A., Torri, L., Casiraghi, M. C., Franzetti, L., Limbo, S., Morandin, F., Quaglia, L., & Pagani,
 312 M. A. (2014). Wheat germ stabilization by heat-treatment or sourdough fermentation: Effects on
 313 dough rheology and bread properties. *LWT-Food Science and Technology*, 59, 1100-1106.

314 Marti, A., Qiu, X. Schoenfuss, T. C., & Seetharaman, K. (2015). Characteristics of Perennial
 315 Wheatgrass (*Thinopyrum intermedium*) and refined wheat flour blends: impact on rheological
 316 properties. *Cereal Chemistry*, 92, 434-440.

317 Marti, A., Augst, E., Cox, S., & Koehler, P. (2015). Correlations between gluten aggregation
 318 properties defined by the GlutoPeak test and content of quality-related protein fractions of winter
 319 wheat flour. *Journal of Cereal Science*, 66, 89-95.

320 Marti, A., Ulrici, A., Foca, G., Quaglia, L., & Pagani, M. A. (2015). Characterization of common
 321 wheat flours (*Triticum aestivum* L.) through multivariate analysis of conventional rheological
 322 parameters and gluten peak test indices. *LWT-Food Science and Technology*, 64, 95-103.

323 Oliveri, P., & Downey, G. (2012). Multivariate class modelling for the verification of food-
 324 authenticity claims. *Trends in Analytical Chemistry*, 35, 74-86.

325 Pagani, M. A., Marti, A., & Bottega, G. (2014). Wheat milling and flour quality evaluation. In:
 326 Bakery Products Science and Technology, Second Edition, eds W. Zhou, Y. H. Hui, I. De Leyn, M. A.
 327 Pagani, C. M. Rosell, J. D. Selman and N. Therdthai. John Wiley & Sons, Ltd, Chichester, UK. pp 17-
 328 53.

329 Pastor, K., Ačanski, M., Vujić, D., Bekavac, G., Milovac, S., & Kravić, S. (2016). Rapid method
 330 for small grain and corn flour authentication using gc/ei–ms and multivariate analysis. Food
 331 Analytical Methods, 9, 443-450.

332 Shepherd, L.V.T., Fraser, P., & Stewart, D. (2011). Metabolomics: A second-generation
 333 platform for crop and food analysis Bioanalysis, 3, 1143-1159.

334 Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., De Jong, S., Lewi, P. J., & Smeyers-
 335 Verbeke, J. (1998). Handbook of Chemometrics and Qualimetrics, Vol. 20B. Elsevier, Amsterdam.

336 Vidya, S., Ravi, R., & Bhattacharya, S. (2013). Effect of thermal treatment on selected cereals
 337 and millets flour doughs and their baking quality. Food and Bioprocess Technology, 6, 1218-1227.

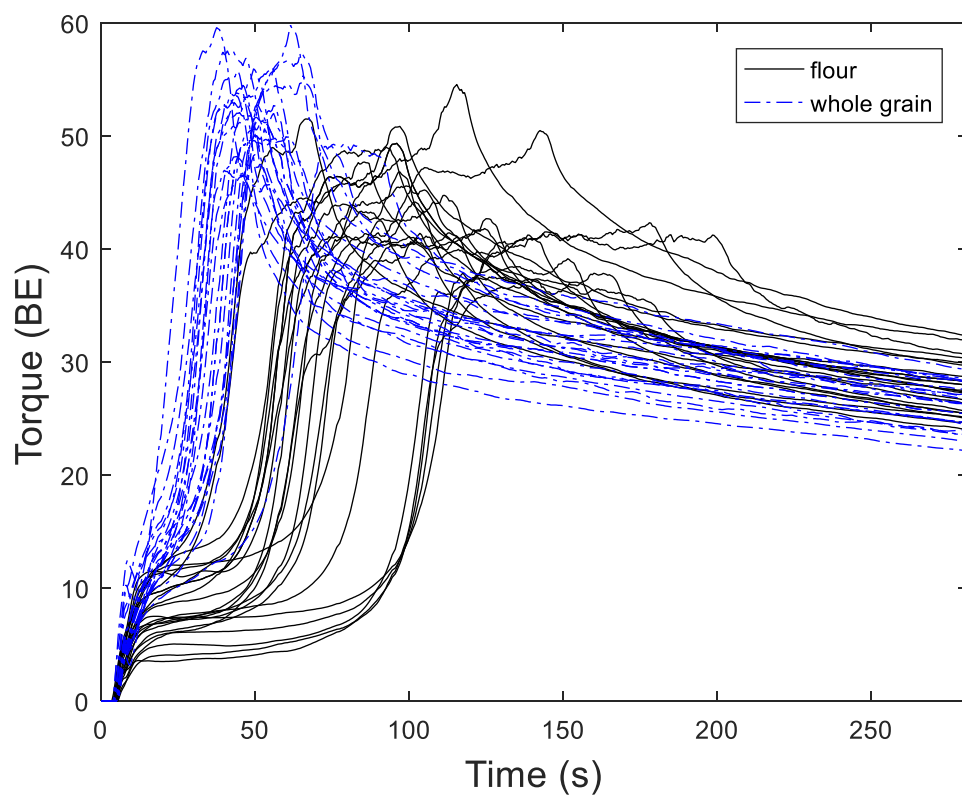
338 **Figure Captions**

339 Figure 1. GlutoPeak profiles of whole grain flours (blue dashed lines) and refined flours (black
340 straight lines).

341 Figure 2. Principal Component Analysis on refined flour data. (a) Score plot and (b) loading plot
342 obtained considering the five most used GlutoPeak indices (PMT, MT, EnMT, EnPM and AgrEn); (c)
343 Score plot and (d) loading plot calculated from the GlutoPeak curves. Samples in both of the score
344 plots are coloured according to a chromatic scale that is proportional to the farinographic stability
345 (FS) value, from blue – for low FS values – to red – for high FS values.

346 AgrEn, Aggregation Energy; EnMT, Energy to Maximum Torque; EnPM, Energy after Maximum
347 Torque; MT, Maximum Torque; PMT, Peak Maximum Time.

348 Figure 3. Principal Component Analysis on whole grain data. (a) Score plot and (b) loading plots
349 obtained considering the the GlutoPeak curves. Samples in the score plot are coloured according to
350 a chromatic scale that is proportional to the farinographic stability (FS) value, from blue – for low FS
351 values – to red – for high FS values.



352

353 Figure 1: GlutoPeak profiles of whole grain flours (blue dashed lines) and refined flours (black straight
 354 lines).

355

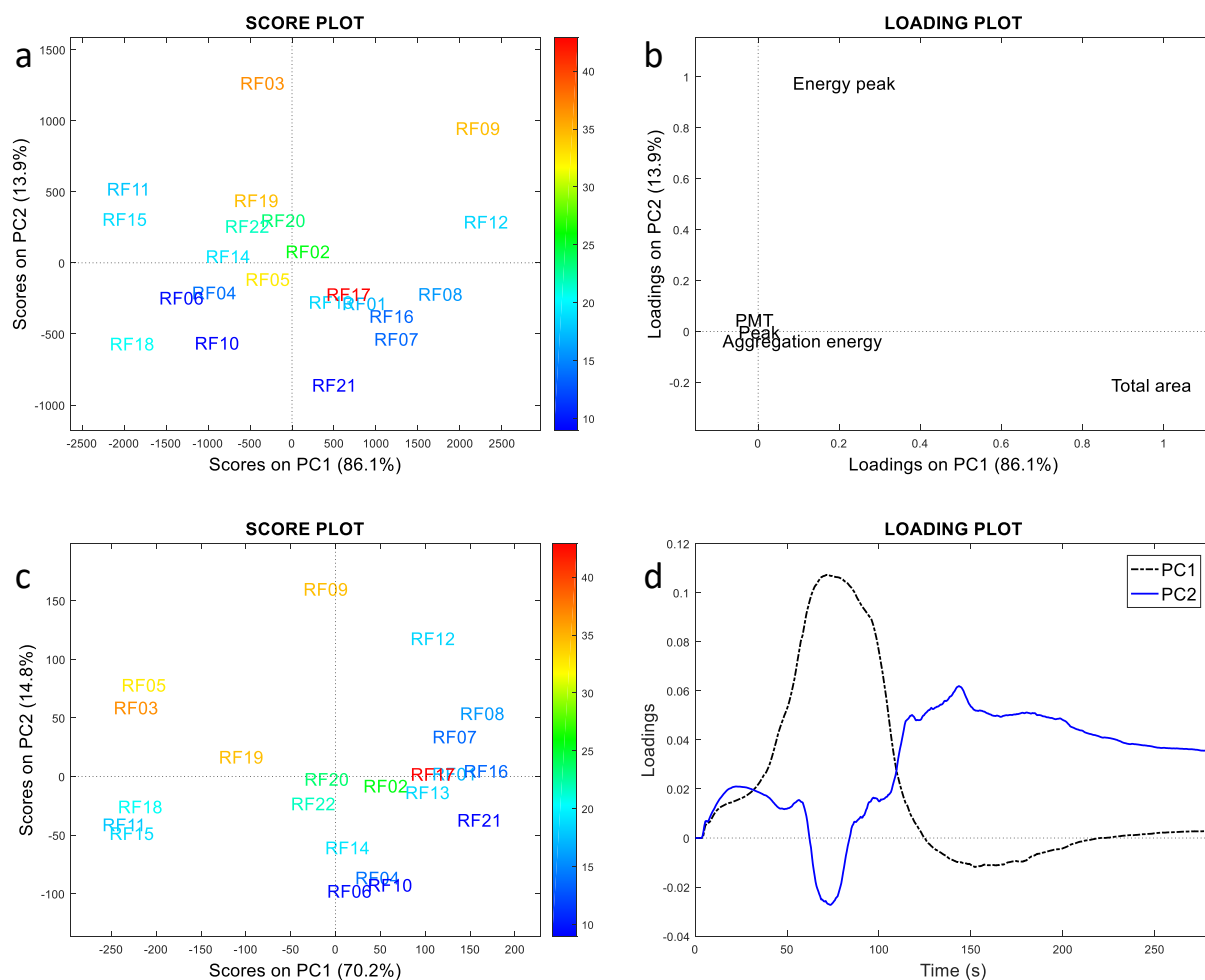
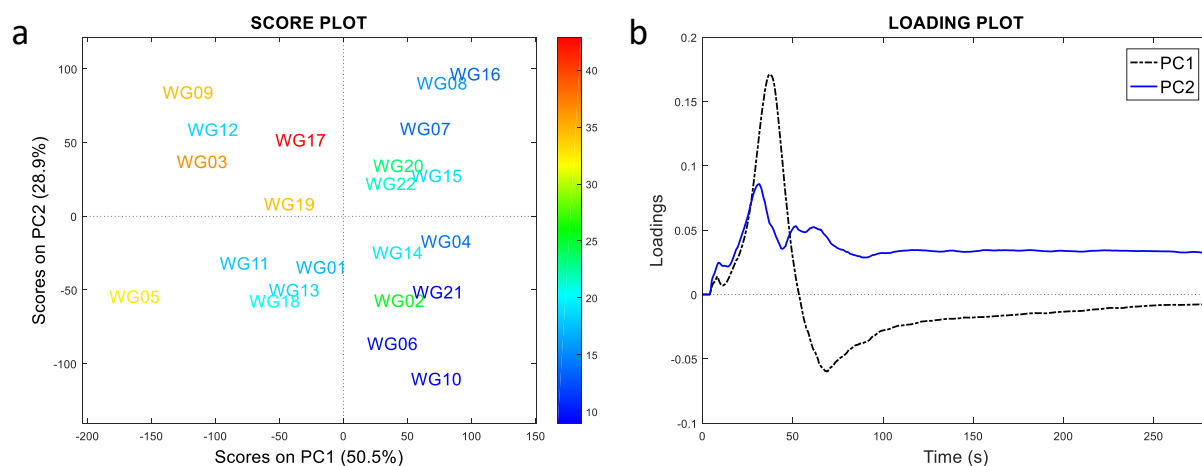


Figure 2. Principal Component Analysis on refined flour data: a) scores plot obtained considering the five most used indices (PMT, MT, EnMT, EnPM and AgrEn); b) loadings plot obtained considering the five most used indices (PMT, MT, EnMT, EnPM and AgrEn); c) and d) scores plot and loadings plot calculated from the GlutoPeak profiles. The samples in both of the scores plot are coloured according to a chromatic scale that is proportional to the FS value, from blue – for low FS value – to red – for high farinographic stability.



373

374

375 Figure 3. Principal Component Analysis on whole grain data: a) scores plot obtained considering the
 376 the GlutoPeak profiles; b) loadings plot calculated from the GlutoPeak profiles. The samples in both
 377 of the scores plot are coloured according to a chromatic scale that is proportional to the FS value,
 378 from blue – for low FS value – to red – for high farinographic stability.

379

380